New product line sales report

Data Validation

The dataset contains 15000 rows and 8 columns before cleaning and validation. All columns were validated against the criteria in the Data Information table.

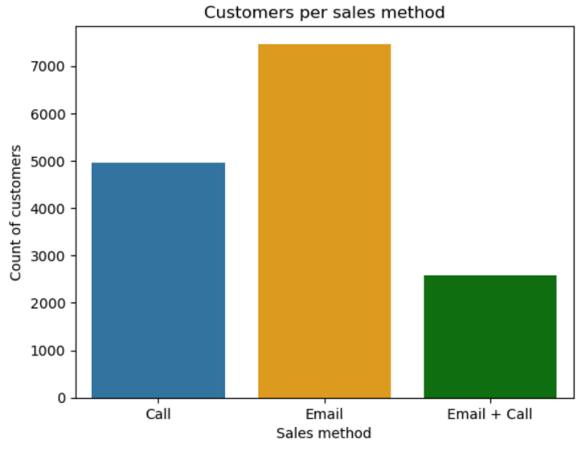
- week: Numeric values without missing values, as described, ranging from 1 to 6. No cleaning was needed.
- sales_method: Converted from object to string type, there were 5 unique values originally used to name the three sales methods. These values were standardized so that only the three sales methods remain. No missing values.
- customer_id: 15000 unique values. Converted from object to string type.
- nb_sold: Numeric values without missing data, as described. No cleaning was needed.
- revenue: numeric values rounded to 2 decimal places, with 1074 missing values. Missing values (NaN) were dropped.
- years_as_customer: Numeric values, with two records above forty (the company's age). These outliers were replaced with the median (3).
- nb_site_visits: numeric values without missing values, as described. Outliers were checked and although 227 records fell below or above the limits based on the IQR method, they were deemed after analysis to be the result of natural variation in the data. Therefore, they were retained without modification.
- state: converted from object to string type, with 50 unique values.

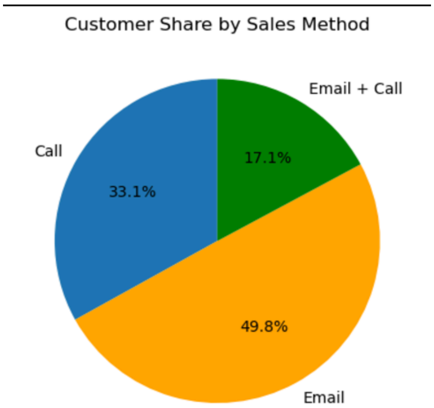
After the data validation, dataset contains 13926 rows and 8 columns without missing values.

Number of customers

To calculate the number of customers for each approach, the dataset before removing null values was used, as missing data only occurred in the revenue variable. This approach ensures a more accurate answer to the sales question.

In the charts below it's seen that almost half of the customers were contacted by Email, followed by Call. Therefore, it is possible to conclude that there is huge room to improve the Email + Call method.



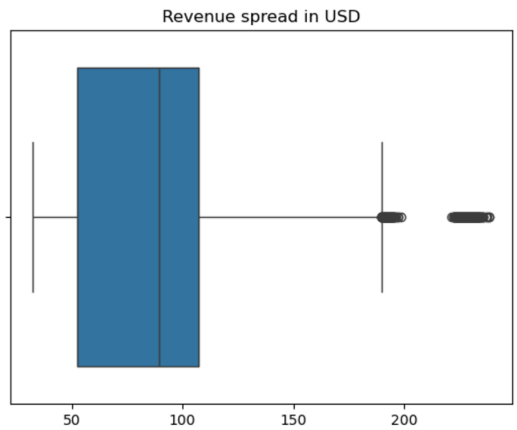


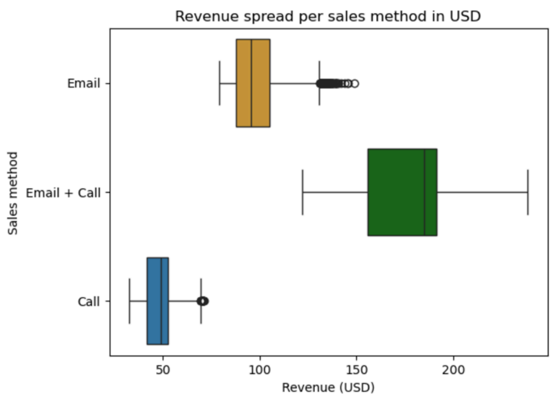
Revenue spread and distribution

First, it's useful to examine the main metrics for the revenue variable, both overall and for each sales method.

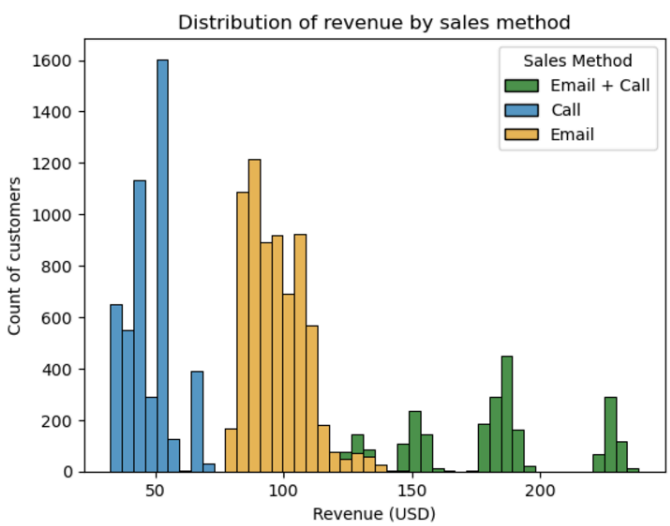
	Overall revenue	Call revenue	Email revenue	Email + Call revenue
Average	$93,93	$47,60	$97,13	$183,65
Stdev	$47,44	$8,61	$11,21	$29,08
Minimum	$32,54	$32,54	$78,83	$122,11
percentile 25	$52,47	$41,47	$87,88	$155,78
percentile 50	$89,50	$49,07	$95,58	$184,74
percentile 75	$107,33	$52,68	$105,17	$191,11
Maximum	$238,32	$71,36	$148,97	$238,32

The following boxplots illustrate how the revenue is distributed overall and across each sales method. It's evident that the Email + Call method outperforms the others in terms of revenue generation.



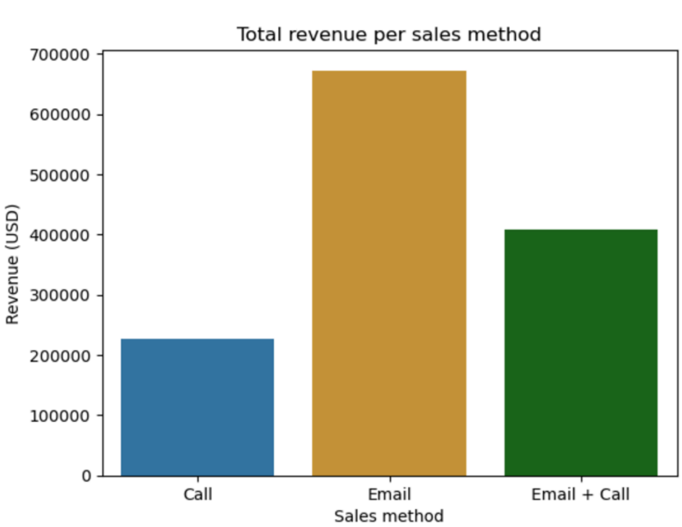


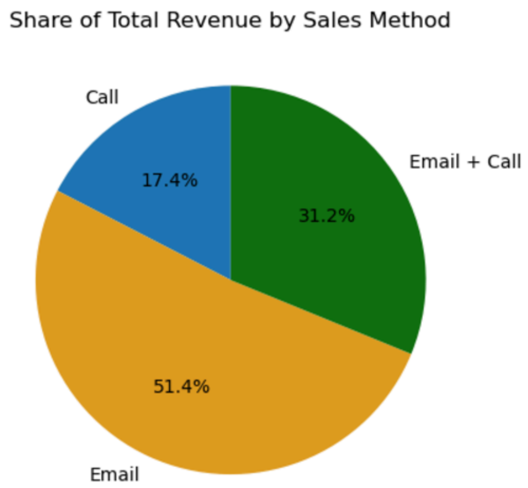
Additionally, it's interesting to explore the superior outliers observed in the overall chart. To analyze these in more detail, a histogram was generated to visualize the data distribution, segmented by sales method.



This histogram clearly demonstrates that each sales method exhibits a distinct revenue distribution. Notably, only the upper bound of the Email method overlaps with the lower bound of the Email + Call method. This finding reinforces the idea that increasing the number of customers in the Email + Call method could significantly boost the company's total revenue.

Expanding on this idea, it’s insightful to analyze how total revenue is composed, not just its spread. A key finding from the analysis emerges when comparing this pie chart with the earlier one showing the percentage of customers by method. The proportions for Calls and Email + Call are almost inverted, highlighting an opportunity to maximize revenue changing the customer distribution among methods.

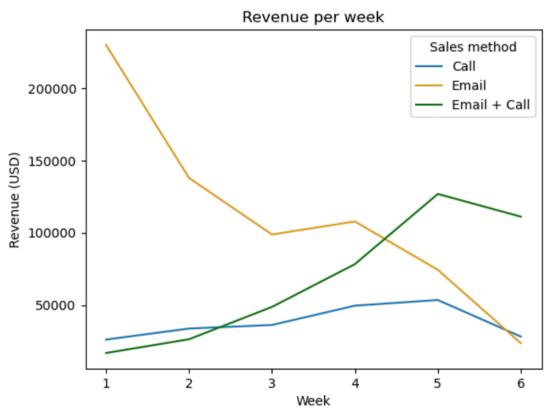


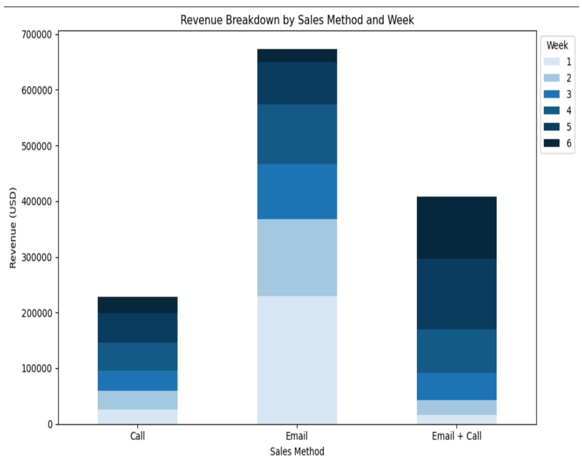


Analysis over time

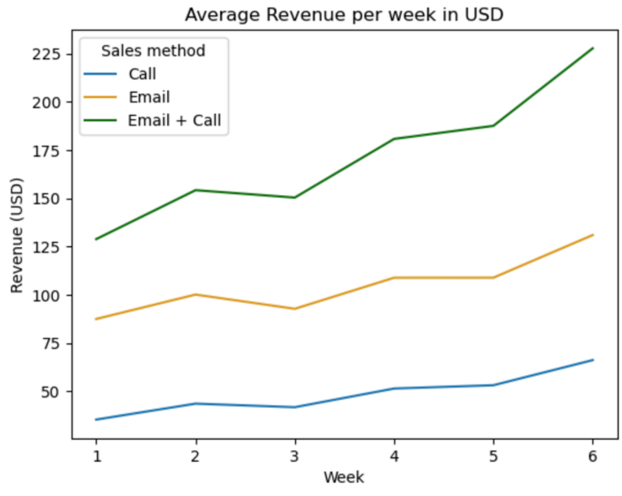
The following two charts illustrate how revenue evolved over time for each sales method. The Email method had an exceptionally strong performance during the first week but declined steadily to become the least revenue-generating method by week six. Calling, on the other hand, maintained a relatively stable revenue level across all weeks.

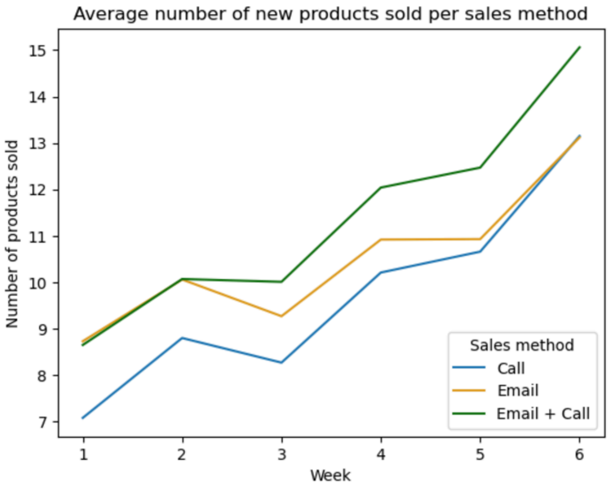
In contrast, the Email + Call method started with minimal revenue in the first two weeks but experienced remarkable growth, ultimately becoming the top-performing method in terms of revenue from week five onward.





Expanding on the time trends analysis, the following charts depict the average performance of each method over time. It becomes evident that all sales methods improve, particularly in the average number of products sold, with the Email + Call method showing the most significant growth. Moreover, the analysis of average revenue highlights the distinct differences between sales methods. The three lines representing the methods never intersect, clearly indicating their varying revenue levels. Notably, the Email + Call method shows the greatest improvement in average revenue.



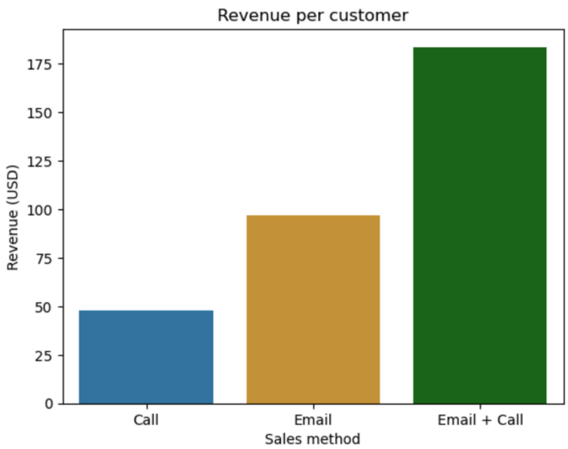


Sales method selection based on the data analysis

Summarizing the findings from the analysis:

- The Email method accounted for nearly half of the sales and generated more than half of the total revenue.
- In contrast, the Call method represented almost a third of the sales but contributed only one-sixth of the total revenue.
- All sales generating $150 or more in revenue were made using the Email + Call method.
- On average, revenue per sale was approximately $50 for Call, $100 for Email, and $180 for Email + Call.

Considering these insights, the **revenue per customer** metric is proposed as a key performance indicator for evaluating the effectiveness of each sales method. This metric is calculated as:
Revenue per customer = Total revenue / Count of customers



The bar chart clearly illustrates that the Email + Call method outperforms the others by a significant margin. Conversely, the Call method has a notably low revenue per customer value, indicating poor efficiency.
Since the Call method requires significantly more time investment, it is recommended to discontinue its use. While it was effective in generating sales, its low revenue per sale and higher time demands make it less viable.

Additionally, it is recommended to focus exclusively on the Email + Call method. It is highly unlikely that following up an Email sale with a Call would negatively impact the sale; on the contrary, it is probable that this additional step would enhance revenue.

Recommendation

Based on the data analysis, the following actions are recommended:

1. Prioritize the Email + Call Method: Discontinue the Calling method as soon as possible, as it is inefficient and generates significantly lower revenue relative to the time investment required.
2. Adopt Revenue per Customer as a Key Metric: Utilize the revenue per customer metric to evaluate and monitor the performance of the new office stationery product line, ensuring that future campaigns focus on maximizing this indicator.
3. Enhance Data Collection for Deeper Insights: Improve the data collection process to include details about customers contacted, not just those who completed purchases. This enhancement would enable more robust analyses of the sales campaign's effectiveness and allow for the calculation of additional performance metrics, such as conversion rate (proportion of sales made from all customers contacted). Also improve the current dataset completing the 1074 missing revenue values.
4. Drive Customer Engagement Through Marketing: Implement targeted marketing campaigns to encourage increased site visits, as evidenced by the heatmap below. This chart segments data into groups based on the number of website visits, divided into intervals (12-14 visits, 15–17 visits, 18–20 visits, etc.). Each cell represents the average revenue generated within a specific range of site visits for each sales method. A clear trend emerges: as the number of visits increases, so does the revenue across all sales methods. This insight highlights the importance of encouraging more frequent visits to the site to boost overall revenue.